

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **2002176619 A**

(43) Date of publication of application: **21.06.02**

(51) Int. Cl.

H04N 5/91
G06T 3/00
G06T 7/60
G10L 11/02
G10L 13/00
H04M 11/00
H04N 1/387

(21) Application number: **2001273890**

(22) Date of filing: **10.09.01**

(30) Priority: **12.09.00 JP 2000275991**

(71) Applicant: **MATSUSHITA ELECTRIC IND CO LTD**

(72) Inventor: **IMAGAWA KAZUYUKI**
TAKADA YUJI
MATSUO HIDEAKI
IWASA KATSUHIRO
YOSHIMURA TETSUYA

(54) MEDIA EDITING METHOD AND APPARATUS THEREOF

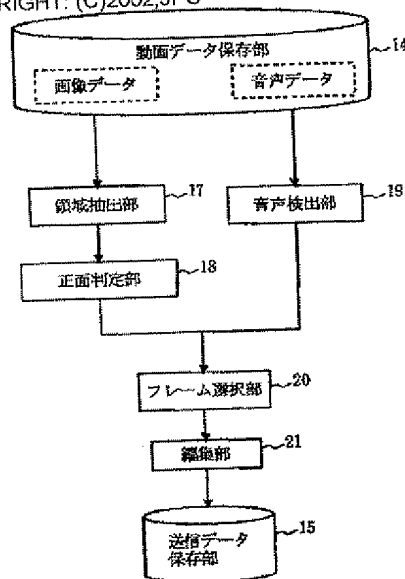
(57) Abstract:

PROBLEM TO BE SOLVED: To provide a media editing apparatus in a communication terminal, such as a mobile terminal that can simply generate media which include messages.

SOLUTION: A moving image data storage section 14 of this media edit device store moving image data that are video-recorded/audio-recorded by a user. An area extracting section 17 extracts an area which includes the user from moving image data. A front discriminating section 18 detects whether the user in the extracted area faces the front. A voice detecting section 19 detects the presence of a voice signal with a prescribed level or higher in the moving image data. A frame selection section 20 decides a start frame and an end frame, on the basis of an output result of the front discrimination section 18 and the voice detection 19. An editing section 21 clips the media based on the decided start and end frames to apply edit, such as image conversion processing, to the clipped media. A transmission data storage section 15 stores the edited

media as transmission data.

COPYRIGHT: (C)2002,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-176619

(P2002-176619A)

(43) 公開日 平成14年6月21日 (2002.6.21)

(51) Int.Cl. ⁷	識別記号	F I	テマコード [*] (参考)
H 0 4 N 5/91		G 0 6 T 3/00	4 0 0 A 5 B 0 5 7
G 0 6 T 3/00	4 0 0	7/60	1 5 0 P 5 C 0 5 3
7/60	1 5 0	H 0 4 M 11/00	3 0 2 5 C 0 7 6
G 1 0 L 11/02		H 0 4 N 1/387	5 D 0 4 5
13/00		5/91	N 5 K 1 0 1
審査請求 未請求 請求項の数21 O L (全 21 頁) 最終頁に続く			

(21) 出願番号 特願2001-273890(P2001-273890)

(22) 出願日 平成13年9月10日 (2001.9.10)

(31) 優先権主張番号 特願2000-275991(P2000-275991)

(32) 優先日 平成12年9月12日 (2000.9.12)

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 今川 和幸

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 高田 雄二

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74) 代理人 100098291

弁理士 小笠原 史朗

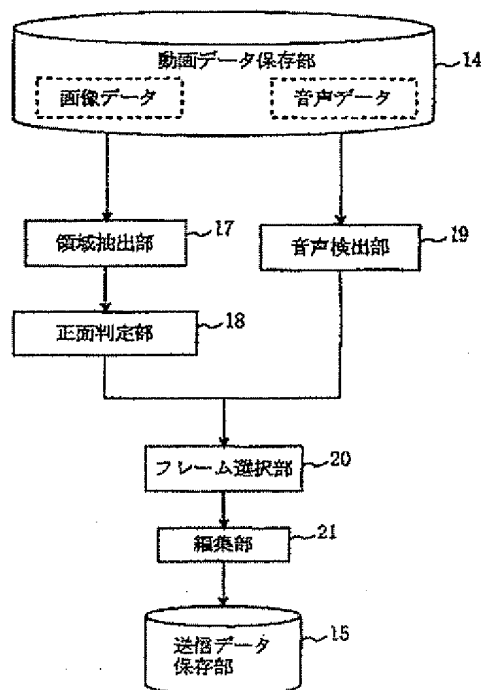
最終頁に続く

(54) 【発明の名称】 メディア編集方法及びその装置

(57) 【要約】

【課題】 携帯端末等のコミュニケーション端末において、メッセージを含むメディアを簡易に作成できるメディア編集装置を提供する。

【解決手段】 本メディア編集装置において、動画データ保存部14は、ユーザが録画・録音した動画データを保存する。領域抽出部17は、当該動画データから、ユーザを含む領域を抽出する。正面判定部18は、抽出された領域内のユーザが正面を向いているか否かを検出する。音声検出部19は、動画データにおける所定レベル以上の音声信号の有無を検出する。フレーム選択部20は、正面判定部18および音声検出部19の出力結果に基づいて、開始および終了フレームを決定する。編集部21は、決定された開始および終了フレームに基づいてメディアをクリッピングして画像変換処理等の編集を行う。送信データ保存部15は、編集されたメディアを送信データとして保存する。



【特許請求の範囲】

【請求項1】 ユーザの一部または全部を被写体として撮影した複数枚の画像から成る画像系列を含むメディアを編集するメディア編集方法であって、前記画像からユーザの一部または全部を含む領域を抽出する領域抽出ステップと、前記領域抽出ステップにおいて抽出された領域に含まれるユーザが予め定められた指定方向を向いているか否かを判定する指定方向顔向き判定ステップと、前記画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に前記指定方向顔向き判定ステップにおいてユーザが前記指定方向を向いていると判定された時点間に対応する画像系列を選択するフレーム選択ステップと、前記フレーム選択ステップによって選択された画像系列を含むメディアを編集する編集ステップとを含む、メディア編集方法。

【請求項2】 前記指定方向顔向き判定ステップは、前記領域抽出ステップにおいて抽出された領域に含まれるユーザが正面を向いているか否かを判定することを特徴とする、請求項1に記載のメディア編集方法。

【請求項3】 前記メディアには音声が含まれており、当該メディアに含まれる音声を検出する音声検出ステップをさらに含み、前記フレーム選択ステップは、前記画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に前記指定方向顔向き判定ステップにおいてユーザが前記指定方向を向いていると判定された時点間であって、かつそれぞれ最初に前記音声検出ステップにおいて音声が検出された時点間の双方に対応する画像系列を選択することを特徴とする、請求項1または請求項2に記載のメディア編集方法。

【請求項4】 前記編集ステップは、前記フレーム選択ステップにおいて選択された画像系列をメタデータ形式の記述によって特定することを特徴とする、請求項1ないし請求項3のいずれかに記載のメディア編集方法。

【請求項5】 前記編集ステップは、前記フレーム選択ステップにおいて選択された画像系列を前記メディアからクリッピングすることを特徴とする、請求項1ないし請求項4のいずれかに記載のメディア編集方法。

【請求項6】 前記編集ステップは、前記フレーム選択ステップにおいて選択された画像系列における最初の画像を初期表示用画像として選択することを特徴とする、請求項1ないし請求項5のいずれかに記載のメディア編集方法。

【請求項7】 前記編集ステップは、前記領域抽出ステップにおいて抽出された領域の位置および大きさに基づいて、前記画像系列において対応する部分領域を算出し、前記部分領域を用いて編集を行うことを特徴とする、請求項1ないし請求項6のいずれかに記載のメディア

編集方法。

【請求項8】 前記編集ステップは、前記部分領域をメタデータ形式の記述によって特定することを特徴とする、請求項7に記載のメディア編集方法。

【請求項9】 前記編集ステップは、前記領域抽出ステップによって抽出された領域に対して全く重ならないかまたは重なる領域がより少なくなるように設定された配置領域に、前記メディアに含まれるテキストを配置することを特徴とする、請求項1ないし請求項8のいずれかに記載のメディア編集方法。

【請求項10】 前記領域抽出ステップは、ユーザの顔を含む顔領域を抽出し、前記編集ステップは、前記領域抽出ステップにおいて抽出された顔領域に対して全く重ならないかまたは重なる領域がより少なくなるように設定された吹き出し領域に、前記メディアに含まれるテキストを配置することを特徴とする、請求項1ないし請求項8のいずれかに記載のメディア編集方法。

【請求項11】 前記編集ステップは、1つ以上の前記画像に対して画像変換処理を行うことを特徴とする、請求項1ないし請求項10のいずれかに記載のメディア編集方法。

【請求項12】 前記領域抽出ステップは、ユーザの顔を含む顔領域を抽出し、前記編集ステップは、前記領域抽出ステップにおいて抽出された顔領域に基づいて算出した顔特徴を参照して、複数のキャラクタ画像および対応する顔特徴を蓄積するキャラクタデータベースから、関連する1つ以上の前記キャラクタ画像を選択することを特徴とする、請求項1ないし請求項11のいずれかに記載のメディア編集方法。

【請求項13】 前記編集ステップは、ユーザから入力された文字データに基づいて、前記顔特徴を算出することを特徴とする、請求項12に記載のメディア編集方法。

【請求項14】 前記編集ステップは、前記領域抽出ステップにおいて抽出された顔領域における顔の縦横比または顔の部位特徴に基づいて前記顔特徴を算出することを特徴とする、請求項12に記載のメディア編集方法。

【請求項15】 ユーザの一部または全部を被写体として撮影した複数枚の画像から成る画像系列を含むメディアを編集するメディア編集装置であって、前記画像からユーザの一部または全部を含む領域を抽出する領域抽出部と、

前記領域抽出部によって抽出された領域に含まれるユーザが予め定められた指定方向を向いているか否かを判定する指定方向顔向き判定部と、

前記画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に前記指定方向顔向き判定部によってユーザが前記指定方向を

向いていると判定された時点間に対応する画像系列を選択するフレーム選択部と、

前記フレーム選択部によって選択された画像系列を含むメディアを編集する編集部とを備える、メディア編集装置。

【請求項16】 前記メディアには音声が含まれており、当該メディアに含まれる音声を検出する音声検出部をさらに備え、

前記フレーム選択部は、前記画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に前記指定方向顔向き判定部によってユーザが前記指定方向を向いていると判定された時点間であって、かつそれぞれ最初に前記音声検出部によって音声が出検された時点間の双方に対応する画像系列を選択することを特徴とする、請求項15に記載のメディア編集装置。

【請求項17】 前記編集部は、前記フレーム選択部によって選択された画像系列における最初の画像を初期表示用画像として選択することを特徴とする、請求項15または請求項16に記載のメディア編集装置。

【請求項18】 前記編集部は、前記領域抽出部によって抽出された領域の位置および大きさに基づいて、前記画像系列によって対応する部分領域を算出し、前記部分領域を用いて編集を行うことを特徴とする、請求項15ないし請求項17のいずれかに記載のメディア編集装置。

【請求項19】 前記編集部は、前記領域抽出部によって抽出された領域に対して全く重ならないかまたは重なる領域がより少なくなるように設定された配置領域に、前記メディアに含まれるテキストを配置することを特徴とする、請求項15ないし請求項18のいずれかに記載のメディア編集装置。

【請求項20】 前記領域抽出部は、ユーザの顔を含む顔領域を抽出し、

前記編集部は、前記領域抽出部によって抽出された顔領域に基づいて算出した顔特徴を参照して、複数のキャラクター画像および対応する顔特徴を蓄積するキャラクターデータベースから、関連する1つ以上の前記キャラクター画像を選択することを特徴とする、請求項15ないし請求項19のいずれかに記載のメディア編集装置。

【請求項21】 ユーザの一部または全部を被写体として撮影した複数枚の画像から成る画像系列を含むメディアを編集するメディア編集装置に、

前記画像からユーザの一部または全部を含む領域を抽出する領域抽出ステップと、

前記領域抽出ステップにおいて抽出された領域に含まれるユーザが予め定められた指定方向を向いているか否かを判定する指定方向顔向き判定ステップと、

前記画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に前記

指定方向顔向き判定ステップにおいてユーザが前記指定方向を向いていると判定された時点間に対応する画像系列を選択するフレーム選択ステップと、

前記フレーム選択ステップによって選択された画像系列を含むメディアを編集する編集ステップとを実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、静止画または動画を含むメディアの編集方法及びその装置に関し、より特定のには、テレビ電話、テレビ会議、ビデオメール、ビデオチャット、ドアフォン（インターフォン）等、少なくとも画像を用いてコミュニケーションを行う場合に、当該画像を含むメディアを編集するためのメディア編集方法及びその装置に関する。

【0002】

【従来の技術】従来より、会議、講演、取材、インタビュー、電話やテレビ電話等を使用した会話などの音声および映像、テレビや監視カメラ等からの映像などを、デジタルディスク、デジタルスチルカメラ、ビデオテープ、半導体メモリなどに記録し、再生する装置が提案されている。このような装置は、記録すべき情報の要点のみを記録者が筆記して記録する方法に比べて、情報を音声や映像の形で漏れなく記録することができるという利点があり、近年では広く用いられている。

【0003】また、近年、ブロードバンド通信が広く利用されるようになったことに伴い、主として個人が、テレビ電話、ドアフォン、またはカメラ付携帯端末等の情報機器を用いて音声や画像を扱い、コミュニケーションを行う場面が増えてきている。例えば、テキストでやりとりされていた従来の電子メールは、音声および動画を用いるビデオメール等へと変容しつつある。また、テレビ電話の普及に伴って、音声での留守録（留守番電話記録）が映像と音声での留守録に変容していくなど、音声と動画を同時に用いてコミュニケーションを行う場面が増えてきている。

【0004】ここで、例えばビデオメールや動画を用いた留守録のように、メッセージなどのデータを一旦蓄積してメディアとして用いる場合には、通常、次のようなステップを経る。

(1)．まず、録画を行う機器の録画開始ボタンを押す。

(2)．次に、伝えたいメッセージを録画する。

(3)．最後に、終了ボタンを押す。

【0005】なお、本発明の各実施形態において、メディアとは、静止画や動画等で構成されたコミュニケーションのためのメッセージ（またはそのデータ）を指すものとする。

【0006】また、蓄積されたメッセージを通信回線を

介して他へ送付する場合には、次のステップを経るのが通常である。

〔4〕、蓄積されたメッセージのどこからどこまでを送付するかを決定し、その部分のみをクリッピングして送付する。

〔0007〕さらに、ビデオメールのようにメッセージを友人等へ送付する場合には、次のステップを経ることもある。

〔5〕、メッセージに対して壁紙などの装飾を施したり、キャラクタなどの絵を添えたり、画像を部分的に切り取るなどのメディア編集を行う。

〔0008〕以上のようなステップのうち、上記〔4〕のステップにおいては、メッセージのどこからどこまでを実際に送付するか（すなわち、クリッピングの開始点および終了点）を決定する際に、ユーザは、蓄積されたメッセージを再生し、確認しながらクリッピングの開始点および終了点を決定しなければならない。しかしながら、カメラ付携帯端末や留守番電話などを用いる場合には、上述のようなメディア編集作業をすることが困難ないし不可能である。

〔0009〕そこで、上記〔3〕のステップを実行した時点で、再度メッセージを再生することなく、送付すべき区間を決定してクリッピングする方法が好ましい。このような方法として、特開平6-343146号公報には、メッセージの記録中にユーザ入力を行い、当該入力が行われたタイミングから一定時間だけ信号を再生する方法が開示されている。しかし、このような方法によれば、記録者が行うユーザ入力のタイミングによって決定される一定時間内の音声や画像だけしかメッセージとして再生することができず、当該一定時間を超えるメッセージは全く再生することができない。また、クリッピングされる区間はユーザ入力のタイミングのみで決定されるので、慎重な操作が要求されるとともに、操作自体も煩わしい。

〔0010〕したがって、上述の従来例のようにユーザ入力を要求することなく、自動的に予め設定された所定の条件に合致する区間を検出して、動画をクリッピングする方法が好ましい。このような方法として、特開平9-294239号公報には、入力音声情報または画像情報のうち、予め設定された所定の条件に合致する区間を検出してクリッピングする手法が開示されている。そして、当該所定の条件としては、所定レベル以上の音声信号の有無、画像の輝度変化や色分布変化、およびカメラ（画像）の動きなどが開示されている。

〔0011〕

【発明が解決しようとする課題】しかし、ほとんど一人の人物がカメラに対面して、メッセージを録画するような通常のビデオメールや動画を用いた留守録の場合には、上述の従来例のような方法を用いることについて、次のような問題点が存在する。

〔0012〕まず、音声信号の有無を条件として動画をクリッピングすることは、ドアホンやテレビ電話での留守録には適しない。なぜなら、誰が訪れてきたか（または誰が電話をかけてきたか）を知りたい場合であっても、音声のメッセージが全く録音されない場合には、クリッピングすることが不可能だからである。また、画像の輝度変化やカメラ（画像）の動きを条件とすることも適しない。なぜなら、通常のビデオメールや動画を用いた留守録の場合には、画像の輝度変化や動きが少ない場合も多く、適切な区間をクリッピングすることが極めて困難だからである。

〔0013〕さらに、上述の従来例では、予め設定された所定の条件に合致する区間を全て検出してクリッピングするが、ビデオメールや動画を用いた留守録の場合には、メッセージの途中で複数回クリッピングされると複数個のメッセージに分解されてしまう。例えば音声信号の有無を条件とした場合に、1つのメッセージの途中でしばらく黙り込み、またメッセージを再開すると、メッセージが複数個に分解されてしまう。しかし、メールや留守録の場合では、これらは1件のメッセージとして構成することが望ましい。また、たとえこれら複数個のメッセージを結合したとしても、1つのメッセージとしては不自然な途切れを含むものになってしまう。

〔0014〕また、上記ステップ〔5〕のように、ビデオメールなどで、装飾を行ったり、表示画面を編集したりする場合、当該操作は煩雑なものになるのが通常である。そのため、現在では、ビデオメールでの装飾や編集は広く行われてはいない。例えば、ゲームセンタなどで見られる娯楽用の自動写真撮影装置のように、専ら静止画に対して壁紙などの装飾を施す形態や、テキストのメールに対して静止画のキャラクタ画像を添付するような形態に留まっている。また、特に市販の携帯電話端末を用いる場合には、親指のみによる操作が望ましいので、装飾や編集の操作はさらに煩雑なものとなりやすい。しかし、携帯端末等の小さな表示画面内に必要な情報を適切に配置するためには、メディア編集処理が特に重要である。にもかかわらず、携帯端末にも適した簡単なメディア編集方法には従来例がない。

〔0015〕そこで、本発明の目的は、ほとんど一人の人物がカメラに対面して録画された動画を含むメディアを適切かつ簡単に編集することができるメディア編集方法を提供することである。

〔0016〕

【課題を解決するための手段および発明の効果】第1の発明は、ユーザの一部または全部を被写体として撮影した複数枚の画像から成る画像系列を含むメディアを編集するメディア編集方法であって、画像からユーザの一部または全部を含む領域を抽出する領域抽出ステップと、領域抽出ステップにおいて抽出された領域に含まれるユーザが指定方向を向いているか否かを判定する指定方向領

向き判定ステップと、画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に指定方向顔向き判定ステップにおいてユーザが指定方向を向いていると判定された時点間に対応する画像系列を選択するフレーム選択ステップと、フレーム選択ステップによって選択された画像系列を含むメディアを編集する編集ステップとを含む。

【0017】上記のように、第1の発明によれば、ユーザが指定方向（例えば、正面や斜め45度方向）を向いているかどうかを判定するため、確実にメッセージを伝える相手に向けられた区間を選択してメディアを編集することが可能になる。

【0018】第2の発明は、第1の発明に従属する発明であって、指定方向顔向き判定ステップは、領域抽出ステップにおいて抽出された領域に含まれるユーザが正面を向いているか否かを判定することの特徴とする。

【0019】上記のように、第2の発明によれば、ユーザが正面を向いているかどうかを判定するため、確実にメッセージを伝える相手に向けられた区間を選択してメディアを編集することが可能になる。

【0020】第3の発明は、第1または第2の発明に従属する発明であって、メディアには音声が含まれており、当該メディアに含まれる音声を検出する音声検出ステップをさらに備え、フレーム選択ステップは、画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に指定方向顔向き判定ステップにおいてユーザが指定方向を向いていると判定された時点間であって、かつそれぞれ最初に音声検出ステップにおいて音声を検出された時点間の双方に対応する画像系列を選択することの特徴とする。

【0021】上記のように、第3の発明によれば、ユーザが指定方向を向いているか否かを判定し、更に音声を検出するために、より確実にメッセージを伝える相手に向けられた区間を選択してメディアを編集することが可能になる。

【0022】第4の発明は、第1ないし第3の発明のいずれかに従属する発明であって、編集ステップは、フレーム選択ステップにおいて選択された画像系列をメタデータ形式の記述によって特定することの特徴とする。

【0023】上記のように、第4の発明によれば、メッセージのデータを切り取ることなく、簡易な方法でメッセージの区間を設定することができ、またメッセージの区間が選択された後であっても、メタデータを編集するだけでメディアを再編集することが可能になる。

【0024】第5の発明は、第1ないし第4の発明のいずれかに従属する発明であって、編集ステップは、フレーム選択ステップにおいて選択された画像系列をメディアからクリッピングすることの特徴とする。

【0025】上記のように、第5の発明によれば、メッセージを伝える相手に向けたメッセージの部分を自動的

にクリッピングすることが可能になる。

【0026】第6の発明は、第1ないし第5の発明のいずれかに従属する発明であって、編集ステップは、フレーム選択ステップにおいて選択された画像系列における最初の画像を初期表示用画像として選択することの特徴とする。

【0027】上記のように、第6の発明によれば、受信中の動画または保存された動画をユーザに対して最初にまず静止画（例えば、プレビュー画面やサムネイル画面など）で示す場合に、画像系列を代表する最適な画面を配置することができる。

【0028】第7の発明は、第1ないし第6の発明のいずれかに従属する発明であって、編集ステップは、領域抽出ステップにおいて抽出された領域の位置および大きさに基づいて、画像系列において対応する部分領域を算出し、部分領域を用いて編集を行うことの特徴とする。

【0029】上記のように、第7の発明によれば、カメラ付携帯端末のように表示画面が小さい場合であっても、ユーザの映る領域のみを効果的に画面に表示することができる。

【0030】第8の発明は、第7の発明に従属する発明であって、編集ステップは、部分領域をメタデータ形式の記述によって特定することの特徴とする。

【0031】上記のように、第8の発明によれば、メッセージのデータを切り取ることなく、簡易な方法でユーザの映る領域を設定することができ、また設定後であっても、メタデータを編集するだけでユーザの映る領域以外の領域を見ることができるようメディアを再編集することが可能になる。

【0032】第9の発明は、第1ないし第8の発明のいずれかに従属する発明であって、編集ステップは、領域抽出ステップによって抽出された領域に対して全く重ならないかまたは重なる領域がより少なくなるように設定された配置領域に、メディアに含まれるテキストを配置することの特徴とする。

【0033】上記のように、第9の発明によれば、カメラ付携帯端末のように表示画面が小さい場合であっても、かつテキストによる情報が存在する場合であっても、画像とテキストとを見易い形で同時に確認できるように配置することができる。

【0034】第10の発明は、第1ないし第8の発明のいずれかに従属する発明であって、領域抽出ステップは、ユーザの顔を含む顔領域を抽出し、編集ステップは、領域抽出ステップにおいて抽出された顔領域に対して全く重ならないかまたは重なる領域がより少なくなるように設定された吹き出し領域に、メディアに含まれるテキストを配置することの特徴とする。

【0035】上記のように、第10の発明によれば、カメラ付携帯端末のように表示画面が小さい場合であっても、かつテキストによる情報が存在する場合であって

も、画像とテキストとを見易くわかりやすい形で同時に確認できるように配置することができる。また、吹き出し形式を用いることによって、通信相手との間のコミュニケーションを活性化することができる。

【0036】第11の発明は、第1ないし第10の発明のいずれかに従属する発明であって、編集ステップは、1つ以上の画像に対して画像変換処理を行うことを特徴とする。

【0037】上記のように、第11の発明によれば、顔画像の解像度や表示色などを受信端末の解像度や表示特性に合わせて画像変換処理を施すことが可能になり、受信端末において見やすい画面を作成することが可能になる。

【0038】第12の発明は、第1ないし第11の発明のいずれかに従属する発明であって、領域抽出ステップは、ユーザの顔を含む顔領域を抽出し、編集ステップは、領域抽出ステップにおいて抽出された顔領域に基づいて算出した顔特徴を参照して、複数のキャラクタ画像および対応する顔特徴を蓄積するキャラクタデータベースから、関連する1つ以上のキャラクタ画像を選択することを特徴とする。

【0039】上記のように、第12の発明によれば、ユーザの顔画像を元に、キャラクタ画像データベースからよく似た特徴のキャラクタを選択することができる。その結果、通信相手との間のコミュニケーションを活性化することができる。

【0040】第13の発明は、第12の発明に従属する発明であって、編集ステップは、ユーザから入力された文字データに基づいて、顔特徴を算出することを特徴とする。

【0041】上記のように、第13の発明によれば、画像の顔特徴をあらわす文字データを利用することによって、ユーザが意図するキャラクタに近いものをキャラクタ画像データベースから選択することが可能になる。

【0042】第14の発明は、第12の発明に従属する発明であって、編集ステップは、領域抽出ステップにおいて抽出された顔領域における顔の縦横比または顔の部位特徴に基づいて顔特徴を算出することを特徴とする。

【0043】上記のように、第14の発明によれば、ユーザの顔画像を元に、キャラクタ画像データベースからよく似た特徴のキャラクタを確実に選択することができる。

【0044】第15の発明は、ユーザの一部または全部を被写体として撮影した複数枚の画像から成る画像系列を含むメディアを編集するメディア編集装置であって、画像からユーザの一部または全部を含む領域を抽出する領域抽出部と、領域抽出部によって抽出された領域に含まれるユーザが指定方向を向いているか否かを判定する指定方向顔向き判定部と、画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、

それぞれ最初に指定方向顔向き判定部によってユーザが指定方向を向いていると判定された時点間に対応する画像系列を選択するフレーム選択部と、フレーム選択部によって選択された画像系列を含むメディアを編集する編集部とを備える。

【0045】第16の発明は、第15の発明に従属する発明であって、メディアには音声が含まれており、当該メディアに含まれる音声を検出する音声検出部をさらに備え、フレーム選択部は、画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に指定方向顔向き判定部によってユーザが指定方向を向いていると判定された時点間であって、かつそれぞれ最初に音声検出部によって音声が出された時点間の双方に対応する画像系列を選択することを特徴とする。

【0046】第17の発明は、第15または第16の発明に従属する発明であって、編集部は、フレーム選択部によって選択された画像系列における最初の画像を初期表示用画像として選択することを特徴とする。

【0047】第18の発明は、第15ないし第17の発明のいずれかに従属する発明であって、編集部は、領域抽出部によって抽出された領域の位置および大きさに基づいて、画像系列によって対応する部分領域を算出し、部分領域を用いて編集を行うことを特徴とする。

【0048】第19の発明は、第15ないし第18の発明のいずれかに従属する発明であって、編集部は、領域抽出部によって抽出された領域に対して全く重ならないかまたは重なる領域がより少なくなるように設定された配置領域に、メディアに含まれるテキストを配置することを特徴とする。

【0049】第20の発明は、第15ないし第19の発明のいずれかに従属する発明であって、領域抽出部は、ユーザの顔を含む顔領域を抽出し、編集部は、領域抽出部によって抽出された顔領域に基づいて算出した顔特徴を参照して、複数のキャラクタ画像および対応する顔特徴を蓄積するキャラクタデータベースから、関連する1つ以上のキャラクタ画像を選択することを特徴とする。

【0050】第21の発明は、ユーザの一部または全部を被写体として撮影した複数枚の画像から成る画像系列を含むメディアを編集するメディア編集装置に、画像からユーザの一部または全部を含む領域を抽出する領域抽出ステップと、領域抽出ステップにおいて抽出された領域に含まれるユーザが指定方向を向いているか否かを判定する指定方向顔向き判定ステップと、画像系列の開始時点から終了時点方向へ、および終了時点から開始時点方向へ向けて、それぞれ最初に指定方向顔向き判定ステップにおいてユーザが指定方向を向いていると判定された時点間に対応する画像系列を選択するフレーム選択ステップと、フレーム選択ステップによって選択された画像系列を含むメディアを編集する編集ステップとを実行さ

せるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【0051】

【発明の実施の形態】（各実施形態の基本的構成）以下、図面を参照して、本発明の各実施形態について概説する。本発明のメディア編集方法及びその装置は、テレビ電話、携帯端末、またはドアフォンのような主として個人または家庭用途向けの画像通信端末を用いて、ユーザがビデオメールを作成し、または訪問者が留守録を作成する際に、使い勝手の良いインタフェースを提供することができる。

【0052】図1は、本発明のメディア編集方法を実現する画像通信を行うメディア編集端末のハードウェア構成を示したブロック図である。図1において、本メディア編集端末は、ユーザに対して情報の入出力を行う入力部1、画像入力部2、画像表示部3、音入力部4、および音出力部5と、入出力される情報を処理する撮像制御部6、音入出力制御部7、表示制御部8、通信部9、記録部10、記録制御部11、信号処理部12、および制御部13とを備える。これらは、システムバスや外部バスなど介して相互に接続されている。なお、これらは、一般的なコンピュータの構成、またはそれに類似する構成である。

【0053】入力部1は、キーボード（テンキー等を含む）やマウス等で構成されており、ユーザがメディアを編集するために必要な各種情報やテキスト形式の文章などをを入力する。

【0054】画像入力部2は、CCDカメラ等で構成されており、少なくとも1枚以上の画像で構成される動画を入力する。なお、画像入力部2は、基本的には後述する音入力部4と連動して動作し、典型的にはメッセージを録画するユーザを撮影した動画画像を取得する。

【0055】音入力部4は、マイクロフォン等で構成されており、装置外部の音声を入力するために用いられる。なお、音入力部4は、基本的に画像入力部2と連動して動作し、典型的にはメッセージを録音するユーザの音声を取得する。

【0056】画像表示部3は、液晶ディスプレイ等で構成されており、ユーザが記録した動画や文字、他人から送信されてきた動画や文字、および操作上必要な各種情報などをユーザに対して表示する。

【0057】音出力部5は、スピーカ等で構成されており、ユーザが記録した音声、他人から送信されてきた音声、および操作上必要な警告音や呼び出し音などをユーザに対して出力する。

【0058】撮像制御部6は、画像入力部2のON/OFFや露出制御など、画像入力部2に対する各種制御を行う。音入出力制御部7は、音入力部4および音出力部5のON/OFFなどの各種制御を行う。表示制御部8は、画像表示部3の制御を行う。

【0059】通信部9は、公衆電話回線や無線などの通信経路を介して、相手の情報処理装置と、後述する実施形態において作成された各種データの送受信を行う。なお、通信部9が用いる通信モードはどのようなものであってもよく、例えば、テレビ電話のような同期型通信でもよいし、メールのような非同期型通信でもよい。

【0060】記録部10は、メモリおよびハードディスク等の記録媒体で構成されており、少なくとも画像入力部2および音入力部4から入力されたデータを記録する。また、記録部10は、CD-ROMやDVDなどの記録媒体と、そのドライブ装置とを含んでもよい。記録制御部11は、記録部10の入出力制御を行う。

【0061】信号処理部12は、デジタルシグナルプロセッサ等で構成されており、画像入力部2から入力された画像信号や音入力部4から入力された音声信号、および記録部10に記録されているデータについて、少なくとも後述する各実施形態における編集において必要な各種処理を行う。

【0062】制御部13は、マイクロコンピュータまたはCPU等で構成されており、上記データの流れを制御し、各種処理を行う。

【0063】なお、本メディア編集端末は、上述した全ての構成部が1つの筐体に収まるような一体型の装置であってもよいし、各構成部がネットワークや信号線を介してデータのやりとりを行う分散型の装置であってもよい。例えば、カメラ付携帯電話端末は、全てが1つの筐体に収められる一体型の装置である。また、ドアフォンは、訪問者とのインターフェースを図るために、少なくとも画像入力部2、音入力部4、および音出力部5がドアの外側近辺に設置されており、それ以外の構成部は居間などに設置された別の筐体に収められるので、分散型の装置である。また、後述するようなキャラクタデータベースを外部に有する分散型の装置も考えられる。

【0064】次に、本発明のメディア編集方法を用いてユーザが送信用データを作成する際の、手順全体の流れについて概説する。図2は、本発明のメディア編集を行う際の情報の流れと処理の手順を示したブロック図である。

【0065】まず、ユーザは、前述した図1の入力部1に対して、画像および音声データの取得を開始するようにコマンドを入力する。その後、ユーザは、画像入力部2および音入力部4を介してメッセージ等を入力して、動画データを作成する。

【0066】通常、作成した動画データには多くの場合、最初と最後の区間にユーザが入力したい情報とは関係ない区間が含まれる。そこで、動画データの最初および最後の不要な区間を省くクリッピング処理を行う。詳しくは後述する。

【0067】次に、データを送信する相手に対して有用な情報（例えば、データの作成日時や作成者の名称）な

どを、作成した動画データと共に一つの画面で表示できるようにレイアウト処理を行う。例えば、クリッピングされた動画データからメッセージを発するユーザを中心にした特定の領域を切り出すようなトリミング処理を行い、トリミングされた領域に対して、さらに素材データとして作成されたテキストやキャラクタ画像などを配置する。なお、素材データとは、画像、テキスト、コンピュータグラフィックス等、動画データに対して付加されるデータ全てを示すものとする。例えば、素材データは、予め作成された画像パターンであってもよいし、文字や記号のパターンであってもよい。また、当該レイアウト処理は、典型的には、MIME (Multipurpose Internet Message (Mail) Extensions)、HTML (Hyper Text Markup Language)、XML (eXtensible Markup Language)、MPEG-7といったメタデータ形式を用いて記述される。以上のような処理を経て、最後に、データを送信する相手方に対するメッセージとして作成された送信データが完成する。

【0068】なお、以上のクリッピング処理およびレイアウト処理は、図1における信号処理部12、制御部13、記録制御部11、および記録部10において行われる。典型的には、これらの処理はコンピュータが実行可能なプログラムによって実現される。当該プログラムは、図示されないCD-ROMや半導体メモリーカードなどのコンピュータが読み取り可能な記録媒体から記録部10等へ入力され、または通信回線を介してダウンロードされる。

【0069】ところで、以上のような処理を行う場合には、従来の方法によれば、ユーザが情報を入力しなければならない場面が多くなることは前述したとおりである。より詳細に説明すれば、クリッピング処理において、ユーザは動画や音声をチェックしながら、クリッピングすべき区間を入力しなければならない。また、レイアウト処理において、ユーザはどのようなレイアウトにすべきかを編集するための操作を行う必要がある。特に、レイアウト処理におけるトリミングでは、動画データの全フレームにわたって切り出すべき領域をユーザが指定しなければならず、非常に煩わしい。また、素材データを付加する場合にも、ユーザは動画データにおける被写体の位置や大きさなどを確認しながら、どこに付加すべきかを入力する必要がある。

【0070】本発明の各実施形態に係るメディア編集装置および方法は、信号処理部12および制御部13の一方または双方において、後述する領域抽出処理、正面判定処理、音声検出処理といった処理を行う。この処理によって、クリッピング処理やレイアウト処理の全てまたは一部において、ユーザが煩わしいと感じる処理（特に、クリッピング処理、トリミング処理、素材配置に関する編集処理）を自動化する気の利いたインタフェースを提供することができる。

【0071】例えば、携帯端末でビデオメールを作成する場合には、メッセージを撮ったらすぐにメールしたいというニーズが高い。そのような場合に気の利いたインタフェースを提供することにより、すぐにメールをする場合であっても、メッセージの最初と最後が予めクリッピングされており、かつユーザが中心にくるようにトリミングされており、場合によって壁紙や吹き出しが配置されており、受信者にとっても見やすいレイアウトとなるように作成されたビデオメールをユーザは簡単な操作（例えば1つのボタンの操作）で作成することができる。なお、以上の処理は本発明の各実施形態において必ずしも全て含まれる必要はなく、用途によって必要なもののみを組み合わせ用いてもよい。以下、本発明の各実施形態について、個別に詳述する。

【0072】（第1の実施形態）第1の実施形態に係るメディア編集装置は、前述した図2のクリッピング処理を自動的に行うことを可能にするものである。図3は、第1の実施形態に係るメディア編集装置における機能構成を示したブロック図である。図3において、本メディア編集装置は、動画データ保存部14と、送信データ保存部15と、領域抽出部17と、正面判定部18と、音声検出部19と、フレーム選択部20と、編集部21とを備える。これらの構成部は、図2のクリッピング処理の一部または全部を実行する。

【0073】動画データ保存部14は、図1の記録部10に対応しており、ユーザがメッセージとして録画・録音した動画データを保存する。領域抽出部17は、保存された動画データから、被写体（ユーザ）の一部もしくは全てを含む特定の領域を抽出する。正面判定部18は、領域抽出部17によって抽出された領域内のユーザが正面を向いているか否かを検出する。音声検出部19は、動画データの各フレームにおいて、所定レベル以上の音声信号の有無を検出する。フレーム選択部20は、正面判定部18および音声検出部19の出力結果に基づいて、開始および終了フレームを決定する。編集部21は、フレーム選択部20によって決定された開始および終了フレームに基づいてメディアをクリッピングし、当該メディアに対して画像変換処理等の編集を行う。送信データ保存部15は、図1の記録部10に対応しており、編集されたメディアを送信データとして保存する。保存されたデータは、適宜送信される。

【0074】次に、これらの構成部の動作を説明する。図4は、或る動画データをクリッピングする処理を説明するための図である。図4において、クリッピング処理は、動画データ保存部14に保存された動画データに対して行われる。動画データには、ユーザの録音したメッセージを含む音声データと、それに同期して録画された画像データとが含まれる。なお、画像データと音声データとは組み合わせられて一つのデータとして構成されてもよいし、画像データと、音声データと、それらの同期関

係が記述されたデータとによって別々に構成されてもよい。図4に示されたデータ例は、ドアホンなどで取得されたメッセージの典型的な例であり、図中のA～Hの8つの場面(時点)が含まれている。これらの時点では、以下の事情が発生している。

- A. データ取得開始
- B. 第1のメッセージ開始
- C. 第1のメッセージ終了
- D. 次のメッセージまでの考慮中の区間
- E. 第2のメッセージ開始
- F. 音声の停止(息継ぎ等)
- G. 第2のメッセージ終了
- H. データ取得終了

【0075】なお、図中のグラフは、経過時間を横軸に、入力された音声レベルを縦軸にとり、A～Hは所定の時刻を示している。また、グラフの上に付された絵は、音声と同時に撮像されたほぼ同時刻の画像データの1シーンであり、例えばドアホンへメッセージを吹き込んでから去って行くまでのユーザを撮像した画像を表している。

【0076】この図4を参照すればわかるように、作成された動画データには多くの場合、最初と最後の部分にユーザが入力したい情報とは関係ない区間が含まれる。なぜなら、一般的に人がメッセージを伝えるときには、メッセージを話し始めるまでの間と、話し終わってからの間が生じるからである。このようなメッセージを含む動画データの特色に着目して、本メディア編集装置は、画像データと音声データに対して、以下のような手法によって、クリッピングすべき区間を自動的に決定する。

【0077】まず、クリッピングの開始時点を検出する手法について説明する。開始時点を検出する場合、まず、開始から終了方向(すなわち、図4のAから見てH方向)にかけて順番に、領域抽出部17は、画像データの各フレームに対して、被写体(ユーザ)の一部もしくは全てを含む領域を検出する。

【0078】なお、従来から、被写体領域を抽出するには、種々の手法が存在する。例えば、特開平5-91407号公報には、動画の隣接するフレーム間の映像信号の相対比較によって動き量の小さい映像部分を背景部分とし、それ以外の部分を被写体領域として抽出する手法が開示されている。また、特開平5-161131号公報には、あらかじめ背景のみの画像を保持しておき、背景のみの画像と動画の各フレームとの差を各画素毎に演算し、差の小さな領域を背景領域とし、差の大きな領域を被写体領域として抽出する手法が開示されている。また、注目領域が被写体の頭部や顔領域である場合には、例えば、画像中の構内領域を検出することにより頭部を検出することができる。このような手法については、谷内らによる「視点固定型パン・チルト・ズームカメラを用いた適応的見え方モデルに基づく人物頭部の検出・追

跡」(MIRU2000 画像の認識・理解シンポジウム講演論文集1 pp. 9-14)に開示されている。また、その他にも色情報により顔を検出する手法、目や口等の顔の一部分に着目する手法、テンプレートマッチングに基づく手法など、周知の様々な手法が存在する。本領域抽出部17は、以上のような従来の手法を適用すれば、被写体領域を容易に抽出することができる。

【0079】次に、領域抽出部17が抽出した領域に対して、正面判定部18は、当該領域に存在するユーザが正面を向いているか否かを検出する。なお、従来から、人物の正面を検出するには、種々の手法が存在する。例えば、正面画像のみをテンプレートとして用意しておき、テンプレートマッチングにより、正面かどうかを検出することができる。また、馬場らによる「Head Classifier: 人物顔画像の実時間分類」(第7回画像センシングシンポジウム講演論文集 PP. 411-416)には、統計的特徴認識手法の一つであるSVM(Support Vector Machine)を用いて、画像中の顔の向きを識別する手法が開示されている。正面判定部18は、以上のような従来の手法を適用すれば、人物が正面を向いているかどうかを判定することができる。また、以上の手法を適用すれば、正面判定部18に代えて指定方向顔向き判定部を設け、画像領域内のユーザが予め指定された所定の方向(例えば、正面に対して右に45度斜めの方向)を向いているか否かを検出するように構成することもできる。そうすれば、斜め方向を向いているユーザの顔などを用いて、予め定められた美しく見える顔の角度の画像を選択することも可能となる。

【0080】一方、音声検出部19は、所定レベル以上の音声信号の有無を条件として、音声の有無を検出する。例えば、当該所定レベルは、周囲の雑音レベルと入力音声の平均レベルとの比較において決定される。また、公知の音声認識等の手法を用いて、人間の音声の有無を検出してよい。

【0081】フレーム選択部20は、正面判定部18および音声検出部19の判定結果を参照して、開始時点から終了時点方向へ各フレームをチェックし、最初に条件を満たした時点のフレームを開始フレームとする。また、終了時点から開始時点方向へ各フレームをチェックし、同様に最初に条件を満たした時点のフレームを終了フレームとする。正面判定部18の判定結果によれば、図4に示された画像中のユーザが最初に正面を向いた時刻は、Bの時点であり、最後に正面を向いた時刻は、Gの時点である。また、音声検出部19の判定結果によれば、最初に入力音声レベルが所定値以上となった時刻は、Bの時点であり、最後の時刻はGの時点である。よって、フレーム選択部20は、共に条件を満たすフレームを選択することにより、開始フレームをB時点のフレームとし、終了フレームをG時点のフレームとする。

【0082】以上のように、本メディア編集装置は、開

始時点から終了時点方向へ、また終了時点から開始時点方向へそれぞれスキャンして最初に条件を満たす時点を選択することにより、D時点付近に存在する次のメッセージまでの渡りの区間をカットすることなく、メッセージ全体をクリッピングすることが可能になる。したがって、ビデオメール等のようにユーザが入力するデータを1件のまとまりとして送信するような用途に適する。

【0083】また、本メディア編集装置は、画像中の被写体の正面判定および音声検出の双方を行うことによって、メッセージとして録画された部分を確実にクリッピングすることができる。例えば、ユーザが正面を向いたまま考え込んでいる場合にも、メッセージを話し始めた時点から確実にクリッピングすることができる。しかし、本メディア編集装置において、音声検出を省略することも可能である。すなわち、ユーザがメッセージを話し始める時には、端末（のカメラ）に正対して行うのが通常であるので、正面判定が特に有効である。また、ユーザがメッセージを吹き込む前の考慮途中に思わず発声してしまった場合には、音声検出が有効でないこともある。したがって、音声検出部19を省略しても、本メディア編集装置は、前述とほぼ同等の効果を有する。

【0084】次に、編集部21は、フレーム選択部20によって判定された開始および終了フレームに基づいて、メディア（動画データ）をクリッピングする。ここで、編集部21は、クリッピングされた区間以外の区間を完全に削除した動画データを作成してもよいし、動画データを削除することなく、クリッピング区間をインデックスとして有するメタデータを作成してもよい。メタデータを作成する場合には、クリッピングされなかった部分に重要な情報が含まれている場合にも、データは存在するのでその部分を利用することが可能となる。以下、メタデータ形式としてMPEG-7形式を用いた場合を例にして説明する。

【0085】メタデータを記述する形式として様々な標準規格、また独自規格が存在するが、最新の標準規格としては、例えば、MPEG-7規格がある。MPEG-7は、Multimedia Content Description Interface (ISO/IEC 15938) と呼ばれるMPEG (Moving Picture Experts Group: ISO/IEC JTC1/SC29/WG11) が制定したMPEG-1、MPEG-2、MPEG-4に次ぐ4番目の標準規格である。この標準規格は、マルチメディア情報の内容を記述するための枠組みを規定しており、デジタルライブラリの検索や編集のためのアプリケーションに用いることを目的とした規格である。また、MPEG-7は、映像・音声を中心とするマルチメディア情報の内容を記述するための記述子 (Descriptor) の標準的な集合を規定する。この記述 (Description) がコンテンツに対して付与されることによ

り、マルチメディア情報の内容に基づいて検索することが可能になる。本規格の実際の記述定義言語は、XML Schema 言語に対して必要な拡張を施したものである。ただし、この拡張についてもXML Schema 言語の文法と互換性が保たれている。

【0086】以上のようなMPEG-7において、コンテンツの特徴を記述するためには、以下の基本要素を組み合わせればよい。

(1) 記述子 (Descriptor: Dとも略称する)

記述子は、マルチメディア・コンテンツにおいて、ある単一の特徴を記述するための基本ツールである。MPEG-7においては、記述子の表記方法 (Syntax) および、意味 (Semantics) が規定されている。

(2) 記述スキーム (Description Schema: DSとも略称する)

記述スキームは、複数の記述ツール間の構造や、意味的關係を規定した枠組みである。MPEG-7においては、同様に記述スキームの表記方法 (Syntax) および、意味 (Semantics) が規定されている。なお、記述スキームを構成する記述ツールとしては、記述子の他、他の記述スキーム自体も含まれる。

(3) 記述定義言語 (Description Definition Language)

記述定義言語は、記述子および記述スキームの表記方法を規定するための言語である。MPEG-7においては、W3C (World Wide Web Consortium) によって標準化されたスキーマ言語である「XML Schema」をベースに、マルチメディア・コンテンツの特徴記述の際に必要な各種データ型などを追加することによって、記述定義言語が規定されている。

【0087】記述スキーム (DS) の例としては、動画データの或る部分へのポイントである「VideoSegmentDS」や、画像内の領域について記述する「StillRegionDS」などがある。また、記述子 (D) の例としては、メディアのフォーマットを記述する「MediaFormat」などがある。なお、MPEG-7においては、標準に定義された記述子および記述スキームの他に、新規の記述子および記述スキームを定義または拡張するための言語 (Description Definition Language) が規定されている。従って、本メディア編集装置においてメタデータを用いる場合には、上記の言語に基づいてメタデータを記述すれば、MPEG-7形式の記述となる。

【0088】図5は、図4のインデックス情報をMPEG-7規格に基づいて記述したメタデータを例示した図である。図5において、本メタデータは、図4のBおよびGの時点について「VideoSegmentDS」

を用いることによって対象となる画像フレームと関連づけている。ここでは、実際の画像フレームとの関連づけは、「MediaTime」の「MediaTimePoint」において行われ、対応するVideoSegmentの時刻が記述される。例えば、図5に示されているような「T13:20:01:1F15」という記述は、「13時20分01秒の第1フレーム目（但し1秒あたり15フレームで0フレームから14フレームまで存在するものとする）」という意味である。このようなメタデータの記述により、動画中の特定の区間のみを選択する再生制御を可能にすることができる。

【0089】なお、以上の記述はあくまでも一例であって、記述形式には制限がない。すなわち、コンテンツと関連づけられる形式で記述されていればよいので、以上の他にあらゆる形式のメタデータで記述されていてもよい。また、メタデータの記述は、図面に表して説明する便宜上、特にテキストフォーマットで示されているが、当該フォーマットには制限がなく、例えばバイナリフォーマットであってもよい。

【0090】以上のように、動画データを削除することなく、クリッピング区間をインデックスとして有するメタデータを作成する場合には、クリッピングされた区間以外の区間を完全に削除した動画データを作成する場合に比べ、自動クリッピング後の再修正を行いたい場合に、メタデータのみを再編集すればよく、自由な編集が可能になる。

【0091】ところで、以上においては、自動クリッピングをするために、フレーム選択部20が出力する開始および終了フレームを利用したが、フレーム選択部20が出力する開始フレームは、編集されたメディアを受信する側の端末画面に最初に表示する画像として定義することもできる。したがって、本メディア編集装置のクリッピング手法はさらに有用である。すなわち、受信中の動画または保存された動画をユーザに対して最初にまず静止画（例えば、プレビュー画面やサムネイル画面など）で示す場合を想定する。なお、このときに用いられる画像を初期表示用画像と定義する。このときに、受信端末の画面には動画の最初のフレーム画像、例えば図4におけるA点の画像が表示される。しかしながら、A点の画像は、送信元のユーザが横を向いて映っており、プレビューやサムネイルなどの初期表示用画像としてはふさわしくない。そこで、図5のようにメタデータを用いて、開始フレームを上記の初期表示用画像として定義する。このことにより、本メディア編集装置は、受信端末に対して新たに初期表示用の静止画を送信することなく、初期表示用画像として適している正面向きのユーザを映したB点のフレーム画像を最初に表示させることができる。なお、新たに初期表示用の静止画を送信する場合には、領域抽出部17および正面判定部18を用いて開始時点から終了時点へ向けてスキャンを行い、B点を

検出する。このB点を初期表示用画像として送信すれば、受信者からは正面に向いたユーザの画像が見えることになる。

【0092】なお、編集部21は、当該初期表示用画像または動画そのものに対して、画像変換処理、例えば解像度の変換処理を行ってもよい。そうすれば、受信端末側の解像度に合わせて最適な表示を行うことができる。解像度を下げる場合には送信時の情報量も少なくすることができる。また、画像変換処理として、2値化処理やグレースケール処理のような表現変換処理を行ってもよい。そうすれば、受信端末側の表示色に合わせて表示を行うことができる。

【0093】以上のことから、本メディア編集装置によれば、ユーザが正面を向いているか否かを判定するため、相手に向けたメッセージの部分を確実に自動的にクリッピングすることができ、受信者は必要な部分だけを再生することが可能となる。また、さらに初期表示用画像としても適した画像を容易に設定することが可能となる。

【0094】（第2の実施形態）第2の実施形態に係るメディア編集装置は、前述したトリミング処理を自動的に行うことにより、受信端末の画面が小さい場合にも見易いレイアウトを効率的に作成することを可能にするものである。

【0095】まず、本実施形態において想定される場面について説明する。通常、ビデオメールなどにより、相手にメディアを送信する場合、当該メディアには、対応する動画データだけでなく、動画の送信者の情報や、件名その他の情報が含まれるのが通常である。図6は、そのようなビデオメールを受信した端末の画面表示例を示した図である。図6に示されるように、表示画面100上には、相手からの動画画面104だけでなく、送信者名、宛先、件名などを含むヘッダ画面101と、テキストによる文章画面102と、適当な装飾等がレイアウトされた装飾画面103とが示されている。

【0096】ここで、図6のような画面を携帯端末のような狭い表示画面で再生する場合には、全体を縮小して表示することが一般的に行われている。しかしながら、狭い表示画面内でさらに全体を縮小した表示を行うと、テキストが見えにくくなったり、動画データの相手の顔が小さくなったりする不都合が生じる。

【0097】また、メッセージとして入力される動画データは、カメラと被写体との位置関係をユーザがあまり気にしないで良いように、通常は、広角のカメラを用いて取得することが多い。そのために、図6の動画画面104のように、ユーザの画像以外の背景領域が多く含まれることになる。したがって、全体を縮小した表示を行うと、さらに相手の顔が小さく見にくくなる不都合が生じる。

【0098】そこで、本実施形態に係るメディア編集装

置は、ユーザを含む部分領域のみを表示画面に表示するようなレイアウト処理を行うために、少なくとも以下のような構成部を備える。なお、第1の実施形態において前述したクリッピング処理を行う場合には、対応する図3の構成部を追加すればよいので、その説明は省略する。

【0099】図7は、第2の実施形態に係るメディア編集装置における機能構成を示したブロック図である。図7において、本メディア編集装置は、動画データ保存部14と、送信データ保存部15と、領域抽出部17と、レイアウト作成部22と、素材データ保存部23を備える。これらの構成部は、図2において前述したレイアウト処理の一部または全部を実行する。

【0100】なお、本メディア編集装置は、第1の実施形態に係るメディア編集装置と共通の構成および動作を有するので、共通する構成部については同一の符号を付して説明を省略する。ただし、本実施形態においては、音声データが特に必要ないことから、動画データ保存部14は、第1の実施形態と同様の動画データを保存してもよいし、また、音声が含まれない画像データを保存してもよい。

【0101】図7において、素材データ保存部23は、図1の記録部10に対応しており、図6に示されるようなテキストや、装飾用画像データなどの素材データを保存する。レイアウト作成部22は、ユーザの操作によって素材データ保存部23から素材データを適宜読み出し、トリミング処理を含むレイアウト処理を行う。詳しくは後述する。

【0102】図8は、トリミング処理例と表示画面例とを示した図である。図8において、図の上部には図6において前述した相手からの動画画面104が示されている。当該画像は、前述した理由によって、ユーザの領域以外の背景領域が多く含まれた画像になっている。そこで、以下のようにユーザ領域のみに対してトリミングを行い、レイアウトを作成する。

【0103】まず、領域抽出部17は、動画データ保存部14に保存された動画の各フレーム画像からユーザの一部（例えば、顔の部分）またはは全部を含む領域を抽出する。このような領域抽出部17の動作については前述した手法によって容易に実現することができる。なお、当該抽出領域はどのような形状であってもよい。

【0104】次に、レイアウト作成部22は、領域抽出部17が抽出する領域に基づいて、動画データのうちの表示しようとする部分領域を算出する。図8において、部分領域は、動画画面104内の太字の四角形で示されている。

【0105】さらに、レイアウト作成部22は、この部分領域に対応する画像と、あらかじめユーザが設定した素材データ（テキストや画像等）を組み合わせるようにして、レイアウトを作成する。図8において、作成され

た表示画面200には、部分領域に対応する動画画像204の他、図6と同様に、ヘッダ画面201と、文章画面202と、装飾画面203とが示されている。このように、レイアウトを決定する際に自動的に動画データが小さな部分領域の画像となることによって、小さな表示画面においても見易い表示が可能になる。

【0106】ここで、レイアウト作成部22は、動画データと素材データのレイアウトを決定するメタデータを作成するのが通常である。したがって、レイアウト作成部22が設定した部分領域についても、同様のメタデータ形式であった方が取り扱いやすく、都合がよい。

【0107】図9は、そのような部分領域についてのメタデータの記述例を示した図である。図9の記述は、第1の実施形態と同様に、MPEG-7形式によって記述されている。ここでは、第1の実施形態において前述した「VideoSegmentDS」を各フレームに適用し、各フレームそれぞれにおける部分領域を「StillRegionDS」を用いて設定するように記述されている。その場合、部分領域情報については、「ContourShape」を用いて部分領域である矩形形状（図中では頂点数は4）および矩形座標（図示されていない）を記述している。

【0108】このようにメタデータを用いる場合には、動画データから部分領域を切り出して新たな動画データを作成する場合のように動画データの容量が小さくなることはない。しかし、受信端末のユーザは、部分領域と全体領域とを好みに応じて切り換えて表示をしたり、他の部分領域を表示するように設定することも可能になり、受信する端末の画面の大きさや受信者の好みなどに合わせてレイアウトを変更することが自由にできることになる。もっとも、この場合においても、レイアウト作成部22が設定した部分領域が初期的に設定されていることは、メッセージの送信者を含む領域が最初に表示されるために都合が良い。

【0109】なお、MPEG-7によれば、図9のように毎フレーム毎に「StillRegionDS」を設定する方法のみならず、動いている領域の情報「MovingRegionDS」や、音と統合した領域情報「AudioVisualRegionDS」等を用いることもできる。これらを包含する基本定義としてマルチメディアコンテンツの一部を示す「SegmentDS」という定義があるが、この定義にもとづいたDSであれば、図9に示す記述と同等の記述をより少ない量の記述で行うことが可能である。

【0110】以上のように、本実施形態のメディア編集装置は、部分的な表示領域を決定することができるので、カメラ付携帯端末のように表示画面が小さい場合であっても、被写体のみの領域を見易い形で画面内に表示することができる。さらに、メタデータの記述に従ってレイアウトを決定する場合には、カメラ付携帯端末やP

C端末のように表示画面の大きさが異なる場合においても、受信者が最適な形で画面に表示することができる。

【0111】（第3の実施形態）第3の実施形態に係るメディア編集装置は、第2の実施形態のトリミング処理とは異なっており、動画の表示領域がなるべく多くなるようにしながら、かつ必要なテキスト表示も行われるようなレイアウトを作成することを可能にするものである。

【0112】まず、本実施形態において想定される場面について説明する。例えば、前述した図6の表示画面100を、なるべく動画画面104の部分が大きくなるようにトリミングして、小さな画面（例えば、携帯電話の表示画面）に表示する場合について考える。ここで、当該小さな画面上に、最低限の情報として「件名」と「本文」と動画とを表示したいとする。しかしながら、実際には、小さな画面には、動画の全領域を表示することしかできず、動画とは別に件名と本文とを表示するスペースは生じない。図10は、このように動画の部分しか表示できない画面表示例を示した図である。

【0113】そこで、本メディア編集装置は、第2の実施形態に係るメディア編集装置と同様の構成ながら、領域抽出部17およびレイアウト作成部22の動作について変更を加え、上記のテキスト情報を表示する。すなわち、領域抽出部17が検出したユーザを含む画像領域に対して、レイアウト作成部22は、「件名」や「本文」などのテキスト情報が重ならないように、または少なくとも重なる度合いが小さくなるように配置を行う。以下、当該動作について詳述する。

【0114】まず、領域抽出部17は、動画データからユーザを含む画像領域を検出し、当該領域の位置および大きさを算出する。次に、レイアウト作成部22は、領域抽出部17が算出した当該領域の位置および大きさと、素材データ保存部23に保存された「件名」や「本文」などの素材データとを受け取る。さらに、レイアウト作成部22は、上述の画像領域に重ならない（または重なる度合いの小さい）領域に対して、素材データを配置する領域を設定する。図11は、ユーザを含む画像領域に重ならない領域に対して、件名を配置した表示例である。図11に例示されるように、テキストの件名は、ユーザの頭の上のスペースに配置されて重ならない。そのため、動画の表示領域が大きいにもかかわらず、必要なテキスト表示も行われるようなレイアウトを作成することができる。

【0115】また、レイアウト作成部22は、素材データを配置する領域を任意の形状として設定してもよい。これを吹き出し領域と定義する。典型的には、吹き出し領域には枠が付され、背景色（例えば白色）を有する。レイアウト作成部22は、この吹き出し領域内に素材データの内容である本文を書き込む。図12は、ユーザを含む画像領域に重なる度合いの小さい領域に対して、本文を配置した表示例である。図12に例示されるよう

に、テキストの本文は、ユーザの左側のスペースに配置されてほぼ重ならない。そのため、動画の表示領域が大きいにもかかわらず、必要なテキスト表示も行われるようなレイアウトを作成することができる。

【0116】なお、図12に示された吹き出し領域の形状には、漫画などで常用されるように、ユーザの口付近に鋭角の突起部分を有する。この突起部分の位置は、画像認識処理によって算出される。具体的には、領域抽出部17は、ユーザを含む領域における口領域を抽出し、その位置を算出する。レイアウト作成部22は、算出された当該位置（またはその付近の適切な位置）に上述の突起部分を配置して、テキストの文字数等を参照しながら、上述の画像領域に重ならない（または重なる度合いの小さい）領域に対して、吹き出し領域を設定する。

【0117】以上のレイアウトされた画面は、受信者の端末の最初の画像（前述の初期表示用画像）として画面上に表示されることが好ましい。すなわち、受信者が受信したメールをあけたときに、最初に図11や図12のような画面を表示する。そして、受信者は、件名や本文の内容のみを確認する。ここで本文の内容が1ページで収まりきらない場合などは、例えば、スクロール処理などを行う。このように、受信者は、最初の表示画面においてのみ本文等の確認を行い、動画再生中は行わない。しかし、もちろん動画再生中にも本文や件名などがスーパーインポーズされた形で再生され、実際の動画メッセージを聞きながら確認することができるように構成してもよい。

【0118】なお、表示されるテキストの内容は件名や本文には限られない。また、まず最初に、図11のような画面が表示され、次に受信者の操作によって図12のような画面が順次表示されてもよいし、これらが同時に合成表示されてもよい。このように、ユーザを含む画像領域に重ならない（または重なる度合いの小さい）領域に対して配置すれば、どのような構成であってもよい。

【0119】以上のように、本メディア編集装置は、動画とテキストとが混在した表示画面においても、受信者が混同せず、また見易く内容を確認することができるようなレイアウトを作成することができる。さらに、吹き出し形式を用いることによって、あたかもテキストの内容を画像中のユーザが話しているような画面となっており、通信相手との間のコミュニケーションを活性化することができる。

【0120】次に、レイアウト作成部22は、以上のレイアウト処理、すなわち、動画に対してテキストを書き込む処理を行うために、第1および第2の実施形態と同様に、レイアウトを決定するメタデータを作成することが好ましい。

【0121】図13は、そのようなレイアウト処理に関するメタデータの記述例を示した図である。図13の記述は、第1および第2の実施形態と同様に、MPEG-

7形式によって記述されており、「MediaDuration」の値、すなわち、メディアの所定のポイントで表示される長さに従って、「Text」タグで囲まれた文面がスーパーインポーズの形式で表示される。このように、メタデータによって記述すれば、実際にテキストを動画に埋め込む処理を行うことなく、テキストを表示することが可能になる。

【0122】（第4の実施形態）第4の実施形態に係るメディア編集装置は、領域抽出部17によって顔領域を抽出し、当該顔領域に対応するキャラクタ画像を選択することによって、ユーザが楽しく表現できるようにメッセージ表現のバリエーションを増大させて、コミュニケーションの円滑化を促進させることを可能にするものである。

【0123】まず、本実施形態において想定される場面について説明する。前述のように、画像を利用したメールでは、装飾を付加することによって楽しさが向上したメールを作成することができる。特に、ゲームセンタにおける自動写真撮影装置のように、本人の画像に対して漫画や3Dグラフィクスなどで構成されたキャラクタを付加することは、特に楽しさや親しみやすさを向上させる効果を有する。図14は、そのようにキャラクタが付加されたビデオメールの受信端末における表示画面例を示した図である。図14に示されるように、表示画面400上には、送信者名、宛先、件名などを含むヘッダ画面401と、テキストによる文章画面402と、相手からの動画画面404とに加えて、さらにキャラクタ画面403とが示されている。以上のようにビデオメールの受信画面を構成すれば、受信者は当該メールに対して親しみを増すと言える。

【0124】さらに、ユーザは、このようなキャラクタを選択する際に、動画の画像や内容と関係のないキャラクタを付加するよりも、画像に関係するキャラクタを選択したいという要求をもつことが多い。本実施形態に係るメディア編集装置は、表示画面の画像が顔画像である場合に、レイアウト処理において、当該顔画像に対応したキャラクタを選択するものである。なお、以下では、以上のようにキャラクタを付したメールを「キャラクタメール」と称する。

【0125】図15は、第4の実施形態に係るメディア編集装置の機能構成を示したブロック図である。図15において、本メディア編集装置は、動画データ保存部14と、送信データ保存部15と、領域抽出部17と、正面判定部18と、編集部26と、キャラクタ選択部24と、キャラクタデータベース25を備える。これらの構成部は、図2において前述したレイアウト処理の一部または全部を実行する。

【0126】なお、本メディア編集装置は、第1の実施形態に係るメディア編集装置と共通の構成および動作を有するので、共通する構成部については同一の符号を付

して説明を省略する。ただし、本実施形態においては、音声データが特に必要ないことから、動画データ保存部14は、第1の実施形態と同様の動画データを保存してもよいし、また、音声が含まれない画像データを保存していてもよい。また、第1の実施形態において前述したクリッピング処理を行う場合には、対応する図3の構成部を追加すればよいので、その説明は省略する。さらに、正面判定部18は、指定方向顔向き判定部であってもよいことは前述したとおりである。

【0127】次に、本メディア編集装置の動作について説明する。領域抽出部17および正面判定部18は、第1の実施形態の場合と同様に動作し、動画中の画像がユーザの正面画像か否かを判定し、編集部26へ入力する。編集部26は、正面画像と判定された画像をキャラクタ選択部24へ入力する。キャラクタ選択部24は、当該画像に基づいて、各種キャラクタをデータベース化して保存するキャラクタデータベース25から、1つまたは複数の候補キャラクタを選択し、選択されたキャラクタに対応するキャラクタIDを編集部26へ入力する。

【0128】ここで、キャラクタ選択部24は、キャラクタデータベース25に保存されている数多くのキャラクタから、キャラクタ候補を選択する。本実施形態においては、前述の正面画像中の顔における顔特徴を抽出し、キャラクタデータベース25に保存されている1つまたは複数のキャラクタを選択する。

【0129】すなわち、キャラクタデータベース25には、2次元のキャラクタ画像データやコンピュータグラフィックスの手法を用いて作成される3次元キャラクタの構成データなどのキャラクタ画像と、当該キャラクタに対応する顔特徴と、対応するキャラクタIDとが、予めキャラクタ情報として登録されている。キャラクタ選択部24は、編集部26から入力された正面画像中の顔における顔特徴を参照して、当該顔特徴に一致するかまたは類似する顔特徴を有する1つ以上のキャラクタ画像をキャラクタデータベース25から選択する。また、上述の顔特徴としては、顔サイズ、顔の縦横比、顔の部位特徴などの特徴量が考えられる。さらに、顔の部位特徴としては、例えば目鼻口サイズ、目鼻口位置関係、髪の毛の量や色などの特徴量が考えられる。また、眼鏡の有無なども顔特徴として利用できる。

【0130】さらに、顔特徴量について詳述する。図16は、頭髮の特徴を用いた顔特徴量について例示的に説明した図である。図16には、6人の異なるユーザの撮影画像と、対応する画像処理結果と、対応する特徴量と、対応する特徴表現とが示されている。なお、撮影画像は写真画像と同様であるが、図を見やすくするために似顔絵として描かれている。また、顔特徴としては、特徴量および特徴表現が例示されているが、これらの一方のみであってもよいし、他の値または表現が用いられて

もよい。

【0131】図16において、処理結果は、頭髮が黒いことを前提として、撮影画像の黒領域を抽出する処理を行った結果を示している。もちろん、頭髮が黒以外の色であっても、頭髮色と同色の領域を抽出することによって同様に処理することができる。また、特徴量としては、正規化面積および正規化周囲長の2つが例示されている。正規化面積とは、頭髮の面積を顔の面積で正規化した値である。正規化周囲長とは、頭髮の周囲長を顔の周囲長で正規化した値である。さらに、特徴表現例としては、髪の色および髪型の2つが例示されている。髪の色とは、一般的な髪の色と平均値と比較して分類したカテゴリ表現であって、具体的には上述の正規化面積を用いて、平均値よりも小さい場合には髪の色が少ないと表現され、平均値よりも大きい場合には髪の色が多いと表現される。同様に、髪型とは、一般的な髪型と比較して分類したカテゴリ表現であって、具体的には上述の正規化周囲長を用いて、平均値よりも小さい場合には髪型が短髪と表現され、平均値よりも大きい場合には髪型が長髪と表現される。このようにして抽出された顔特徴量またはその特徴表現を用いれば、類似する1つ以上のキャラクター画像をキャラクターデータベース25から選択することができる。

【0132】なお、このような顔特徴量の抽出手法としては、従来より種々の技術が開示されている。例えば、山口らによる「顔向きや表情の変化にロバストな顔認識システム“Smart face”」（電子情報通信学会論文誌 Vol. J84-D-II, No. 6）には、部分空間法を用いて顔領域を検出し、分離度フィルタにより目・鼻・口などの顔部品を検出する手法が開示されている。本メディア編集装置は、各種周知の技術の少なくとも1つを用いることにより、容易に顔特徴量を抽出することができる。

【0133】また、抽出された顔特徴量を参照して、キャラクターデータベース25に登録された各キャラクター候補を選択するためには、前述した特徴表現を用いてもよいし、登録された顔特徴量との相関値を計算してもよい。ここでは、候補とするのにふさわしいものとして設定された抽出閾値を計算された相関値が超えた場合には、当該キャラクターを候補として抽出する。キャラクター選択部24は、抽出された候補キャラクターに対応するキャラクターIDを編集部26へ通知する。

【0134】編集部26は、通知されたキャラクターIDに基づいて、候補とされたキャラクター画像をユーザへ表示する。図17は、使用するキャラクターを選択する編集画面を例示した図である。図17において、3つの候補キャラクターが表示されており、矢印はユーザが選択しようとするキャラクターを示すカーソルである。もちろん、カーソルを用いることなく、選択しようとするキャラクターの画像を反転させたり、周囲の枠を太くするなどの手

法を用いてもよい。

【0135】図17のような編集画面をもとに、ユーザは使用するキャラクターを選択する。編集部26は、選択されたキャラクターIDを記述したメタデータを作成するメディア編集を行って、送信データを作成する。もちろん、キャラクター画像自体を送信データに組み込んでもよい。こうして作成された送信データは、送信データ保存部15に保存されて、適宜のタイミングで受信者に宛てて送信される。

【0136】図18は、上記のようにして作成された送信データを受信した場合における受信端末画面を例示した図である。図18に示されるように、受信端末画面の左下には、送信者であるユーザが選択したキャラクターが表示され、右下には動画メッセージが表示される。

【0137】図19は、上記送信データを受信した場合における受信端末画面の別例を示した図である。図19に示されるように、受信端末画面の下側には、送信者であるユーザが選択したキャラクターが表示されている。そして、例えば動画メッセージが再生される際には、当該キャラクターは表示されず、その表示位置に動画を表示してもよい。このようなレイアウトは、編集部26によって作成されてもよいし、受信側で設定されてもよい。

【0138】なお、候補となるキャラクターは1つであってもよく、その場合、候補選択を行うことなく、より簡単にメールを作成することが可能になる。

【0139】また、編集部26は、キャラクター選択部24に対して、ユーザが入力した特徴量を示す文字列（前述の特徴表現など）を通知してもよい。例えば、ユーザは「髪の色が多く、短髪である」といった文字列を入力する。そうすれば、キャラクター選択部24は、キャラクターデータベース25に保存されている図16に示されるような特徴表現と照合して、候補キャラクターを選択することができる。このように、文字列によって特徴量を示すことによって、本メディア編集装置は、ユーザの意図するキャラクターにさらに近い候補キャラクターを選択することが可能になる。

【0140】さらに、前述したように、本メディア編集装置は、全ての構成部が1つの筐体に収まるような一体型の装置のみならず、各構成部がネットワークや通信線を介してデータのやりとりを行う分散型の装置であってもよい。したがって、例えば、キャラクター選択部24およびキャラクターデータベース25がネットワークを介してアクセス可能であるような、本メディア編集装置とは別の場所に存在している形態の場合が考えられる。図20は、このような分散型のメディア編集装置ないしシステムの構成を示したブロック図である。

【0141】図20において、このような分散型メディア編集装置は、第1ないし第3の実施形態に係るメディア編集装置の機能の一部または全てを含むキャラクターメール編集端末501と、キャラクターメール編集端末50

1とは別の場所に設けられたキャラクタ選択部724と、キャラクタデータベース725とを含み、これらはネットワーク600を介して相互に接続されている。このような分散型のメディア編集装置も、前述の一体型のメディア編集装置と同様の構成を有し、動作を行うので、同様の効果を得ることができる。さらに、図20の分散型メディア編集装置は、キャラクタ選択部724およびキャラクタデータベース725をキャラクタメール編集端末501以外の端末、例えば、メールを受信するキャラクタメール受信端末502などが用いることができる。すなわち、受信時において、キャラクタメールに含まれるキャラクタIDを受信したキャラクタメール受信端末502は、キャラクタデータベース725からキャラクタ画像を受け取るだけでよいので、端末自体に多くのデータを保持している必要がない。また、キャラクタメール受信端末502が返信時にメディア編集装置として動作する場合でも、キャラクタ選択部724およびキャラクタデータベース725を共有することができる。

【0142】このように、分散型メディア編集装置は、キャラクタ選択部724およびキャラクタデータベース725を複数のユーザによって共に利用することができるので、各端末はこれらの構成部を内蔵する必要がなく、また多くのキャラクタを保存するデータベースを利用することができる。

【0143】以上のように、本メディア編集装置によれば、ユーザは、撮影された動画から抽出された正面画像を用いて多くの登録キャラクタから候補を絞り込み、好みのキャラクタを容易に選択してキャラクタメールを作成することが可能となる。また、このようなキャラクタメールを用いることによって、通信相手との間のコミュニケーションを活性化することができる。

【図面の簡単な説明】

【図1】本発明のメディア編集方法を実現する画像通信を行うことができるメディア編集端末のハードウェア構成を示したブロック図である。

【図2】本発明のメディア編集を行う際の情報の流れと処理の手順を示したブロック図である。

【図3】第1の実施形態に係るメディア編集装置における機能構成を示したブロック図である。

【図4】或る動画データをクリッピングする処理を説明するための図である。

【図5】図4のインデックス情報をMPEG-7規格に基づいて記述したメタデータを例示した図である。

【図6】動画データ、送信者の情報、件名その他の情報が含まれるビデオメールを受信した端末の画面表示例を示した図である。

【図7】第2の実施形態に係るメディア編集装置における機能構成を示したブロック図である。

【図8】トリミング処理例と表示画面例とを示した図で

ある。

【図9】部分領域についてのメタデータの記述例を示した図である。

【図10】動画とは別に件名と本文とを表示するスペースが生じないような画面表示例を示した図である。

【図11】ユーザを含む画像領域に重ならない領域に対して、件名を配置した表示例である。

【図12】ユーザを含む画像領域に重なる度合いの小さい領域に対して、本文を配置した表示例である。

【図13】動画に対してテキストを書き込むレイアウト処理に関するメタデータの記述例を示した図である。

【図14】キャラクタが付加されたビデオメールの受信端末における表示画面例を示した図である。

【図15】第4の実施形態に係るメディア編集装置の機能構成を示したブロック図である。

【図16】頭髪の特徴を用いた顔特徴量について例示的に説明した図である。

【図17】使用するキャラクタを選択する編集画面を例示した図である。

【図18】キャラクタメールを受信した場合における受信端末画面を例示した図である。

【図19】キャラクタメールを受信した場合における受信端末画面の別例を示した図である。

【図20】分散型のメディア編集装置ないしシステムの構成を示したブロック図である。

【符号の説明】

- 2 画像入力部
- 3 画像表示部
- 4 音入力部
- 5 音出力部
- 6 撮像制御部
- 7 音入出力制御部
- 8 表示制御部
- 9 通信部
- 10 記録部
- 11 記録制御部
- 12 信号処理部
- 13 制御部
- 14 動画データ保存部
- 15 送信データ保存部
- 17 領域抽出部
- 18 正面判定部
- 19 音声検出部
- 20 フレーム選択部
- 21 編集部
- 22 レイアウト作成部
- 23 素材データ保存部
- 24 キャラクタ選択部
- 25 キャラクタデータベース
- 26 編集部

- 100, 200, 400 表示画面

101, 201, 401 ヘッダ画面

102, 202, 402 文章画面

103, 203 装飾画面

104, 204, 404 動画画面

403 キャラクタ画面
- 501 キャラクタメール編集端末

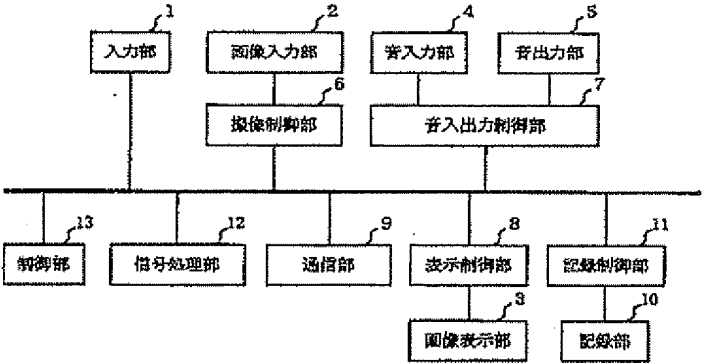
502 キャラクタメール受信端末

600 ネットワーク

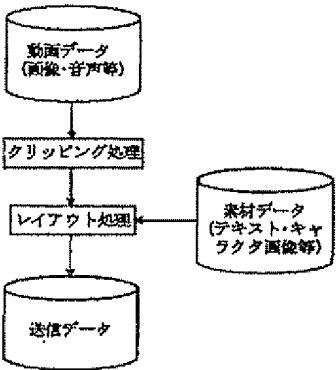
724 キャラクタ選択部

725 キャラクタデータベース

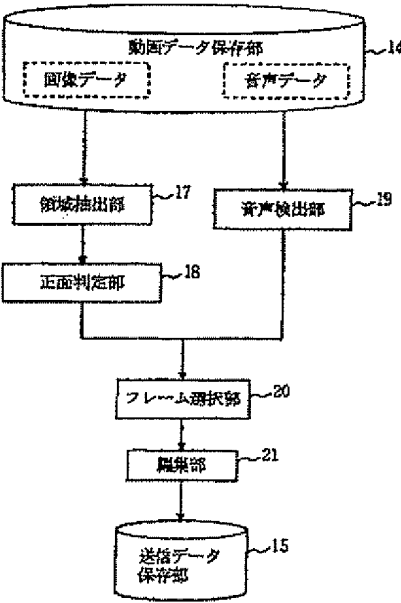
【図1】



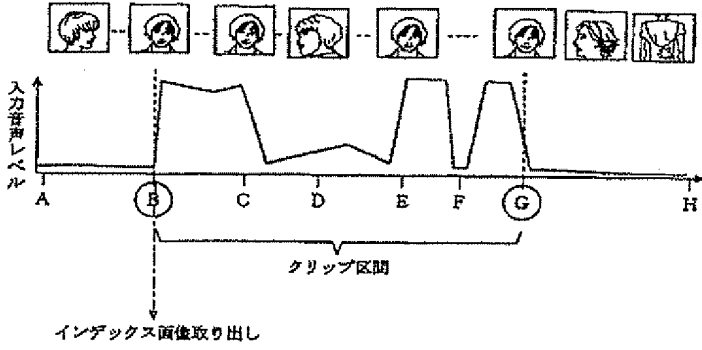
【図2】



【図3】



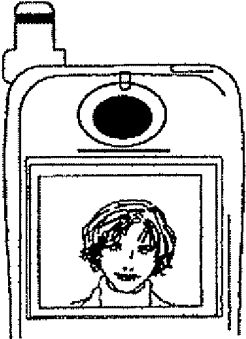
【図4】



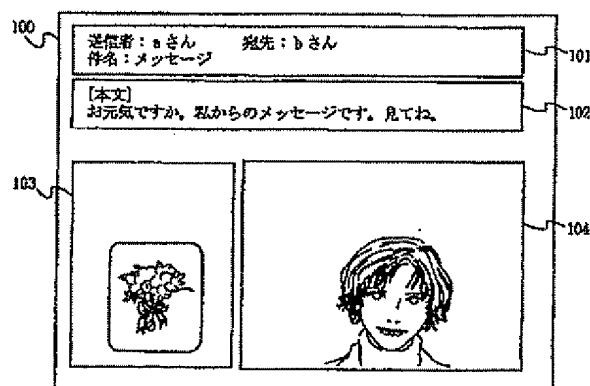
【図5】

```
...
<VideoSegment id="開始点">
  <MediaTime>
    <MediaTimePoint>T13:20:01:1F15</MediaTimePoint>
  </MediaTime>
</VideoSegment>
...
<VideoSegment id="終了点">
  <MediaTime>
    <MediaTimePoint>T13:23:20:15F15</MediaTimePoint>
  </MediaTime>
</VideoSegment>
```

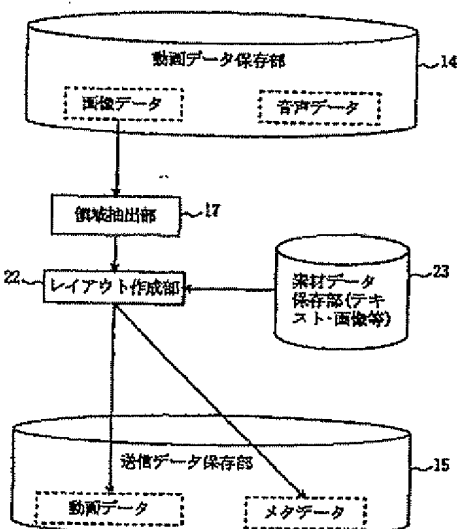
【図10】



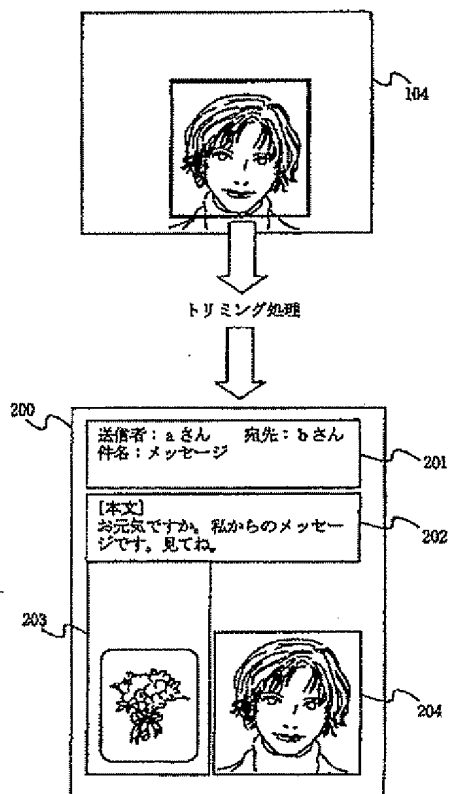
【図 6】



【図 7】



【図 8】



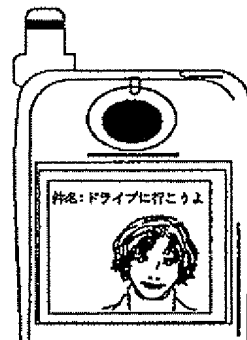
【図 9】

```

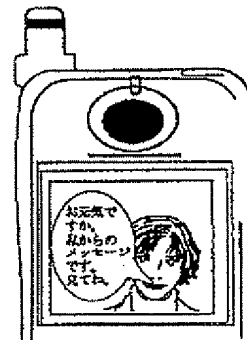
...
<VideoSegment id= "第1フレーム" >
  <MediaTime>
    <MediaTimePoint>T13:20:01:1F15</MediaTimePoint>
  </MediaTime>
  <StillRegion id= "部分領域情報" >
    <ContourShape numberOfPeaks= "4" >
      ...
    </ContourShape>
  </StillRegion>
</VideoSegment>
...
<VideoSegment id= "第2フレーム" >
  <MediaTime>
    <MediaTimePoint>T13:20:15F15</MediaTimePoint>
  </MediaTime>
  <StillRegion id= "部分領域情報" >
    <ContourShape numberOfPeaks= "4" >
      ...
    </ContourShape>
  </StillRegion>
</VideoSegment>
...
<VideoSegment id= "第3フレーム" >
  <MediaTime>
    <MediaTimePoint>T13:29:02:3F15</MediaTimePoint>
  </MediaTime>
  <StillRegion id= "部分領域情報" >
    <ContourShape numberOfPeaks= "4" >
      ...
    </ContourShape>
  </StillRegion>
</VideoSegment>

```

【図 11】



【図 12】



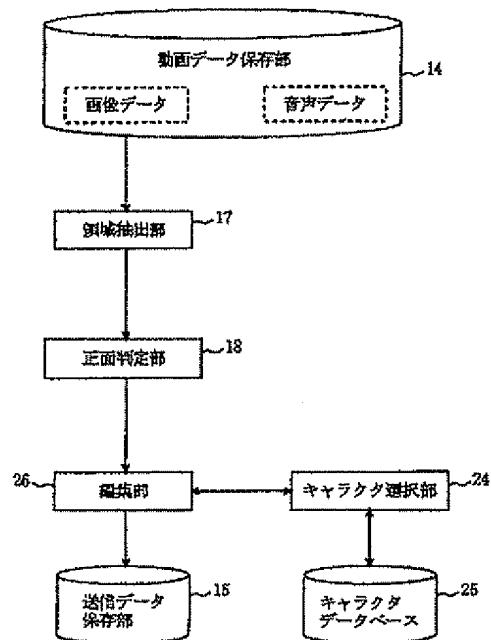
【図13】

```

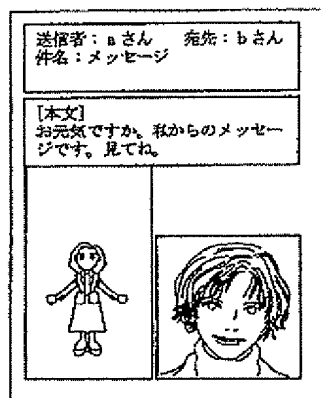
<?xml version="1.0" encoding="UTF-8"?>
<Video id="Video1">
  <VideoText id="VideoText1">
    <textType="Superimposed" FontSize="40" FontType="ゴシック">
      <MediaTimePoint>T13:20:01:1F15</MediaTimePoint>
      <MediaDuration>P765</MediaDuration>
    </MediaTime>
    <Text>
      <Text>お元気ですか。私からのメッセージです。見てね。</Text>
    </Text>
  </VideoText>
</Video>

```

【図15】



【図18】

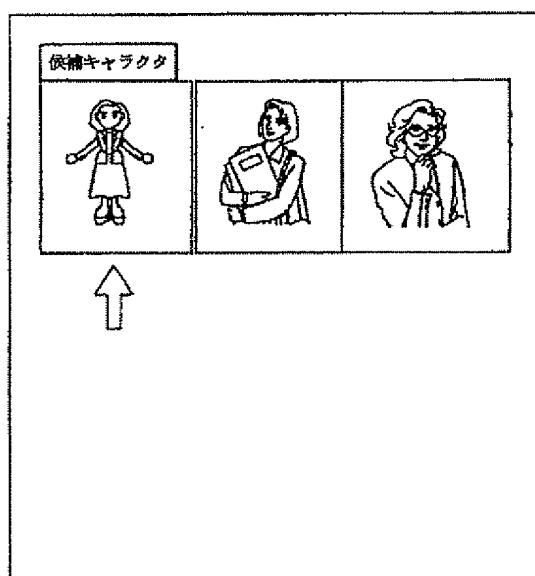


【図14】

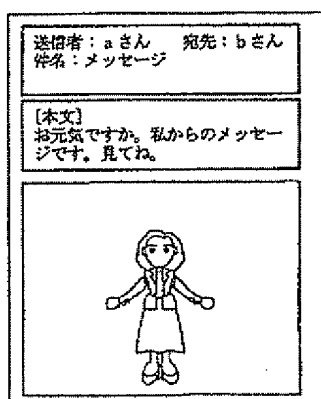


表示画面













【図17】



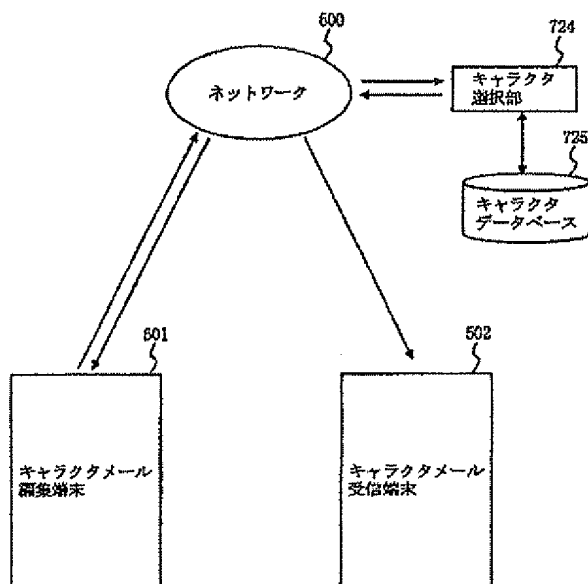
【図19】



【図16】

撮影画像							
処理結果							
特徴量	正規化面積	0.289	0.581	1.777	2.257	2.439	1.82
	正規化周囲長	3.71	9.31	12.3	13.1	24.3	22.7
特徴表現例	髪の色	少ない		多い			
	髪型	短髪				長髪	

【図20】



フロントページの続き

(51) Int. Cl. 7

H 0 4 M 11/00

H 0 4 N 1/387

識別記号

3 0 2

F I

G 1 0 L 3/00

テ-コ-ド (参考)

B 5 L 0 9 6

Q

(72) 発明者 松尾 英明

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 岩佐 克博

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 吉村 哲也
大阪府門真市大字門真1006番地 松下電器
産業株式会社内

Fターム(参考) 5B057 GE09
5C053 FA08 FA14 GB08 GB11 GB19
JA01 LA01 LA14
5C076 AA02 AA14 BA06 CA10
5D045 AB01
5K101 LL12 MM07 NN06 NN07 NN18
NN21
5L096 FA67 JA03 JA11